Supplement: Sample analyses

All analyses were carried out using Crux v2.0. Supplementary files containing sample input files, command lines, and output files exemplifying the analyses described below can be obtained from `http://noble.gs.washington.edu/proj/cruxtoolkit`.

Human variation data set

The human variation data set was derived from lymphoblastoid cell lines from 95 HapMap individuals[1], including 53 Caucasians, 33 Yorubans, 9 eastern Asians and one Japanese. Protein lysates were subjected to detergent cleanup, cysteine alkylation, trypsin digestion and isobaric tandem mass tag (TMT) labeling. Digested peptides were labeled with sixplex TMT, and the six TMT-labeled samples were equally mixed to generate the final digest mixture. All digest mixtures were analyzed on an LTQ Orbitrap Velos (Thermo Scientific) equipped with an online 2D nanoACQUITY UPLC System (Waters). During data acquisition, the full MS scan was performed in the orbitrap in the range of 400–1800 m/z at a resolution of 60000, followed by the selection of the 10 most intense ions for HCD-MS2 fragmentation using a precursor isolation width window of 1.5 m/z. Ions with singly charged state or unassigned charge states were rejected for MS2. Ions within a 10 ppm m/z window around ions selected for MS2 were excluded from further selection for fragmentation for 60 s. The final data set consists of 9,092,380 fragmentation spectra in 560 RAW files (`http://www.peptideatlas.org/PASS/PASS00230`) with a total size of 224 GB.

The IPI.Human database version 3.74 contains 89,599 protein sequences. Peptides were digested using trypsin, including suppression of cleavage by proline, allowing a minimum length of 7 amino acids and up to two missed cleavages. One static modification was included: carbamidomethylation (57.02146 Da) of cysteine. TMT labeling (229.16293 Da) of lysine and N-terminal amino acids were treated as variable modifications. One additional variable modification–oxidation (15.995 Da) on methionine—was included, allowing up to three occurrences per peptide. These rules yielded a total of 19,874,734 distinct peptides.

The spectra were searched with Tide using reverse peptide decoys and 10 ppm precursor monoisotopic mass tolerance. The resulting top-scoring PSMs per spectrum were subjected to analysis using Percolator with default parameters.

Erythrocytic cycle of *Plasmodium falciparum*

The *Plasmodium* data set is derived from a recent study of the erythrocytic cycle of the malaria parasite *Plasmodium falciparum*[2]. *P. falciparum* 3D7 parasites were synchronized and harvested in duplicate at three different time points during the erythrocytic cycle: ring (16 ± 4 h postinvasion), trophozoite (26 ± 4 h postinvasion), and schizont (36 ± 4 h postinvasion). Parasites were lysed, and duplicate samples were reduced, alkylated, digested with Lys-C, and then labeled with one of six TMT isobaric labeling reagents. The resulting peptides were mixed together, then fractionated via strong cation exchange into 20 fractions, desalted and then analyzed via LC-MS/MS on an LTQ-Velos-Orbitrap mass spectrometer. All MS/MS spectra were acquired at high resolution in the Orbitrap. In total, the data set consists of 348,157 spectra, divided into unphosposphorylated (227,875 spectra) and phosphorylated (120,282 spectra) components.

The *Plasmodium* protein database was downloaded from NCBI Entrez on January 8, 2014, by searching for "Plasmodium falciparum 3D7." The resulting database contains 11,737 proteins and 8,669,430 amino acids. Peptides were digested by Lys-C with up to three allowed missed cleavages. Three static modifications were included: carbamidomethylation of cysteine and TMT labeling of lysine and N-terminal amino acids. Variable modifications of up to three sites per peptide were allowed for methionine oxidation (15.9949 Da). For the phosphorlyation spectra, additional variable modifications on serine, theoronine, and tyrosine phosphorylation (79.9663 Da) were also allowed. These digestion rules and variable modifications led to a total of 8,546,260 distinct peptides.

The spectra were searched with Comet using reverse peptide decoys and 50 ppm precursor monoisotopic mass tolerance, allowing for -1, 0, +1, +2, and +3 isotopic offsets. The fragment mass bin size was set to 0.03 Da. The resulting top five PSMs per spectrum were subjected to analysis using Percolator with default parameters.

References

[1] Wu, L. *et al.* Variation and genetic control of protein abundance in humans. *Nature* **499**, 79–82 (2013).

[2] Pease, B. N. *et al.* Global analysis of protein expression and phosphorylation of three stages of *Plasmodium falciparum* intraerythrocytic development. *Journal of Proteome Research* **12**, 4028–4045 (2013).